Izvestiya NAN Armenii. Matematika, vol. 44, no. 2, 2009, pp. 51-58.

TRUTH, BELIEF AND EXPERIENCE – A ROUTE TO INFORMATION

F. TOPSØE

University of Copenhagen, Copenhagen, Denmark
E-mail: *topsoe@math.ku.dk*

ABSTRACT. In gratitude to Klaus Krickeberg who introduced the author to Shannons information theory, this contribution is devoted to certain basic considerations which are consistent with, yet carry you beyond Shannons original ideas from 1948, cf. [13]. Fact is that since Shannons pioneering work – to a great extent centred around the notion of *entropy* – a jungle of alternative entropy measures have been suggested. Philosophical speculation will lead us through this jungle and lay out a narrow path of special entropy measures, the so-called *Tsallis entropies*, thereby providing these entropy measures with special credibility.

Dedicated to the 80th birthday of Klaus Krickeberg

1. INTRODUCTION

Undeniably, modern information theory started sixty years ago with Shannons path-breaking paper [13]. Twelve years later, Krickeberg, then guest professor at the University of Århus, introduced a small group of students, including the author, to the fascinating new world. Emphasis was on interpretations. The students should see that Shannons concepts were "just the right ones" .

Since Shannons pioneering work, a multitude of other concepts, mainly measures of *entropy* and *divergence*, have been suggested and applied in many fields of science. This development is technically intriguing and of some fascination. However, for many of the new concepts there are no convincing arguments to the effect that these concepts too are "just the right ones" to work with.

Different areas of science have different needs. We shall have the needs of statistical physics in mind. A key feature of the all-important notion of what is there often referred to as *Boltzmann-Gibbs-Shannon entropy* is its *additivity* (additivity over independent subsystems). However, there is a need for other quantities since experimental evidence has shown that for certain phenomena, classical thermodynamics

does not lead to agreement with observation. It was suggested by Tsallis in 1988, cf. [15], that by basing the thermodynamic considerations on a new class of entropies, now known as *Tsallis entropies*, satisfactory agreement with data was possible in cases where the classical theory had failed. The interested reader is referred to the comprehensive documentation in the "Tsallis literature" which can be traced through the bibliography maintained by Tsallis, cf. [15].

The class of Tsallis entropies, which comprises Shannon entropy as a special case, have received much attention in the statistical physics literature, but also been met with criticism due mainly to a lack of transparent interpretations. In our approach, we shall focus on three concepts, *truth* held by "nature" , *belief* as expressed by man and *experience* acquired through observation by man. Based on the hypothesis that there is a functional relationship between the three concepts and on a natural variational principle, classical- as well as non-classical measures of entropy and other essential quantities are derived. The approach aims at a genuine interpretation, rather than relying on formal mathematical analogies or on axiomatic characterizations.

Our approach is philosophical or speculative, if you wish, and we shall use the excuse of the "Festschrift atmosphere" to surpass certain technical difficulties. Instead, the contribution is an introduction, with focus on the ideas. It will be followed-up by a more comprehensive and technical publication. As it is, the contribution is an appetizer which may, so is the intention, be enjoyed by the Festschrift readers.

2. CONTEMPLATING

Let us put ourselves in the shoes of the physicist who is planning to set-up *experiments* and to engage in associated *observations*. Borrowing terminology from philosophy, the physicist operates in a certain *world* and is interested in studying particular *situations* from this world. The physicist might argue as follows:

1: I find that *truth, belief* and *experience* are concepts of key importance on the way to *information*. I seek the truth, am restricted in my planning of experiments by my beliefs and after observation, I will know by experience through the data observed how truth manifests itself to me. Thinking about it, I ask why should not what I see in terms of data depend not only on truth but also on belief? Accepting this idea, I introduce a functional relationship $z = \Pi(x, y)$. Here, x, y and z represent, respectively a *truth instance*, a *belief instance* and an *experience-* or *data instance*. These *instances* are objects associated with any particular situation I may be interested in.

The function Π is the *global interactor*. It is a characteristic of the world of which I am a part.

As extreme examples, I point to the *classical world* where experience is a genuine reflection of truth. This world is characterized by the global interactor $\Pi(x, y) = x$. And, as another extreme, I can think of, I mention a *black hole* characterized by the global interactor $\Pi(x, y) = y$. In such a world, I can only get out what I myself put in.

2: I am interested in many quite different situations and in my overall planning I will, just as did Shannon, focus on concepts which are independent of semantic content. Therefore, I apply probabilistic reasoning across semantic differences. In this way I will also enable quantitative reasoning. Thus, instances x, y and z related to truth, belief and experience in a specific situation will be probability vectors $(x_i)_{i\in\mathbb{A}}$, $(y_i)_{i\in\mathbb{A}}$ and $(z_i)_{i\in\mathbb{A}}$ with $\mathbb{A}$, the *alphabet*, a set of *indices* which identify the various basic *events* associated with the situation in question. I will concentrate on discrete distributions. To me, they are the more fundamental ones.

I assume that the global interactor acts *locally*, i.e. is of the form $\Pi(x, y) = (\pi(x_i, y_i))_{i\in\mathbb{A}}$ for some real valued function π defined on $[0, 1] \times [0, 1]$. This function is the *local interactor* or just the *interactor*.

As examples, the local interactor corresponding to the classical world is the projection $(x, y) \mapsto x$ on the first coordinate whereas the local interactor corresponding to a black hole is the projection $(x, y) \mapsto y$ on the second coordinate.

3: I must be prepared for other forms of interaction than those connected with either a classical world or a black hole, but will always assume that the interactor is *sound*, i.e. that $\pi(x, x) = x$ for all $x \in [0, 1]$. Stronger conditions should be considered and in this connection, it appears sensible to impose conditions of *consistency*: I will call the interactor *weakly consistent* if, for any pair (x, y) of probability vectors, $x = (x_i)_{i\in\mathbb{A}}$ and $y = (y_i)_{i\in\mathbb{A}}$, $\sum_{i\in\mathbb{A}} z_i = 1$ with $z_i = \pi(x_i, y_i)$ for $i \in \mathbb{A}$. If, with the same assumptions on x and y, it can be concluded that $z = (z_i)_{i\in\mathbb{A}}$ is in fact a probability distribution, I will say that π is *strongly consistent*.

4: Any event I may observe entails a certain *effort* on my part. This effort I shall also refer to as the *local description cost*. Before setting up *experiments*, I will determine the effort I am willing to or have to devote to any event I may be faced with. It can only depend on the assigned belief-value y_i and is denoted $\kappa(y_i)$. The function

$\kappa : y \mapsto \kappa(y)$, defined on $[0,1]$ and with values in $[0,\infty]$, I refer to as the *descriptor*. As 1 represents *certainty*, $\kappa(1) = 0$ must hold. I do not want to distinguish between descriptors that only differ by a scalar factor, and therefore introduce an assumption of *normalization*. As $\kappa(1) = 0$ and as I do not want to assume that $\kappa(0)$ is finite, I impose the condition $\kappa'(1) = -1$ as the natural normalization condition.

5: I will apply a *principle of separability* and consider my *total effort* related to observations in a given situation to be the sum of local efforts associated with the basic events. In so doing, I must take into account the weights with which I will experience the various basic events. The total effort I also refer to as the *total description cost* or simply the *description cost*. This cost, denoted by the letter Φ, is thus the weighted sum of individual contributions, i.e., with x for truth- and y for belief instances,

$$\Phi(x,y) = \sum_{i\in\mathbb{A}} \pi(x_i,y_i)\kappa(y_i)\,. \tag{2.1}$$

6: I will attempt to minimize description cost and shall appeal to the variational principle that the smallest value is obtained when there is a *perfect match* between truth and belief, i.e. when $y = x$. This principle I call the *perfect match principle*. The quantity

$$\sum_{i\in\mathbb{A}} \pi(x_i,y_i)\kappa(y_i) - \sum_{i\in\mathbb{A}} x_i\kappa(x_i) \tag{2.2}$$

represents my *frustration*, as it compares the actual description cost with the smallest possible cost, had I only known the truth. The perfect match principle may, therefore, also be formulated by saying that frustration is the least, in fact disappears, when $y = x$.

Theoretically, if I knew $x = (x_i)_{i\in\mathbb{A}}$, *minimal description cost* is what I would aim at. It is an important quantity. In anticipation, I call it *entropy* and denote it by the letter H:

$$H(x) = \inf_{y=(y_i)_{i\in\mathbb{A}}} \Phi(x,y) = \sum_{i\in\mathbb{A}} x_i\kappa(x_i)\,.^{1} \tag{2.3}$$

The quantity (2.2) also appears important. It is tempting to call it *"frustration"* but, again in anticipation, I better call it *divergence*. I shall denote it by the letter D:

$$D(x,y) = \Phi(x,y) - H(x)\,. \tag{2.4}$$

[1] In order to allow a singular case – the case $q = 0$ of Theorem 1 below – to fit into the framework, the infimum should be restricted to run over probability distributions y with a support which contains the support of x.

I may leave it to further technical discussion how divergence should be defined in cases with infinite entropy – or I may neglect this possibility as it cannot represent any physical reality.

3. CONCLUDING

Theorem 1. *With assumptions and definitions as introduced above, assuming only that the interactor is weakly consistent, the number $q = \pi(1,0)$ must be non-negative and, to each $q \in [0,\infty[$, there is only one interactor and one descriptor which fulfill the conditions imposed. These functions, denoted by π_q and κ_q, are determined by the formulas*

$$\pi_q(x,y) = qx + (1-q)y\,, \tag{3.1}$$

$$\kappa_q(y) = \ln_q \frac{1}{y}\,, \tag{3.2}$$

where the q-logarithm is given by

$$\ln_q x = \begin{cases} \ln x \ \textit{if } q = 1, \\ \frac{x^{1-q}-1}{1-q} \ \textit{if } q \neq 1\,. \end{cases} \tag{3.3}$$

It is assumed implicitly that the interactor and the descriptor satisfy suitable regularity conditions related to continuity and differentiability.

Regarding the proof, we shall here only give a brief outline: The formula (3.1) is readily derived from the assumption of weak consistency. Then, the only possible form for the descriptor, (3.2), follows from pretty standard variational arguments. Indeed, introducing Lagrange multipliers, one is soon led to the differential equation

$$(1-q)\kappa(x) + x\kappa'(x) = -1\,, \tag{3.4}$$

and (3.2) follows in view of the normalization condition $\kappa'(1) = -1$. The final step of the proof, that with (3.1) and (3.2) the perfect match principle holds, follows from (3.10) below or, alternatively, one may observe the close tie to entropy- and divergence- measures as derived by an approach due to Bregman, cf. the recent papers [14] and [12].

Note that strong consistency holds if and only if $0 \leq q \leq 1$.

The accompanying quantities, description cost, entropy and divergence are denoted Φ_q, H_q and D_q, respectively. They are given through (2.1), (2.3) and (2.4), i.e.

$$\Phi_q(x,y) = \sum_{i\in\mathbb{A}} \pi_q(x_i,y_i)\kappa_q(y_i)\,, \tag{3.5}$$

$$H_q(x) = \sum_{i\in\mathbb{A}} x_i \kappa_q(x_i)\,, \tag{3.6}$$

$$D_q(x,y) = \sum_{i\in\mathbb{A}} \Big(\pi_q(x_i,y_i)\kappa_q(y_i) - x_i\kappa_q(x_i)\Big)\,. \tag{3.7}$$

For $q = 1$ we obtain Shannon type quantities: *Kerridge inaccuracy*, *Shannon entropy* and *Kullback-Leibler divergence*, cf. [9] and the standard reference [4]. For $q \neq 1$, the formulas above may be written in a number of ways. The following forms are useful:

$$\Phi_q(x,y) = \sum_{i\in\mathbb{A}} \Big(\frac{q}{1-q} x_i y_i^{q-1} + y_i^q - \frac{1}{1-q} x_i\Big)\,, \tag{3.8}$$

$$H_q(x) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} (x_i^q - x_i) = \frac{1}{1-q}\Big(\sum_{i\in\mathbb{A}} x_i^q - 1\Big)\,, \tag{3.9}$$

$$D_q(x,y) = \sum_{i\in\mathbb{A}} \Big(\frac{q}{1-q} x_i y_i^{q-1} + y_i^q - \frac{1}{1-q} x_i^q\Big)\,. \tag{3.10}$$

In (3.8) the linearity in x is evident. This is important as it leads to a relatively easy approach to key optimization problems. For an indication of this, see [14] and [12]. In (3.9) we recognize the family of *Tsallis entropies*, cf. [15]. Note the special case $q = 0$ corresponding to a black hole where the entropy only depends on the number n of elements in the support of x, indeed, $H_0(x) = n - 1$. In (3.10) the main convenience of the formula is due to the fact that the summands are non-negative. This can be exploited to give an easy proof of the "q-version" of the *fundamental inequality* of information theory: $D_q(x, y \geq 0$ with equality if and only if $x = y$. This is valid for any $q > 0$. For $q = 0$, one finds that $D_0 \equiv 0$. The formula (3.10)also points to a possible extension to go beyond the case of discrete distributions.

The general formulas (2.1), (2.3) and (2.4) indicate that for the determination of the quantities involved one needs to know the interactor π as well as the descriptor κ. Two facts – to be discussed more thoroughly in a planned publication – should be emphasized. Firstly, through the perfect match principle, the descriptor is uniquely determined from the interactor. Therefore, in principle, only the interactor needs to be known. Secondly, different interactors may well determine the same descriptor. Thus, knowing only the descriptor, you cannot determine divergence or description cost. But you *can* determine the entropy function.

It is instructive to consider the family $(\kappa_q)_{0\leq q<\infty}$ of descriptors. This is a descending family of decreasing functions on $[0, 1]$. The largest descriptor, $\kappa_0(x) = \frac{1}{x} - 1$, is associated with a black hole. For $0 \leq q \leq 1$, the descriptors are convex and assume the value ∞ for $x = 0$. For $q = 1$, we find the descriptor $\kappa_1(x) = \ln\frac{1}{x}$ associated

with the classical world. Then, for $1 < q < 2$ the descriptors are convex and finite valued, also for $x = 0$. The special descriptor $\kappa_2(x) = 1 - x$ is affine. For $2 < q < \infty$ we find descriptors which are concave with $\kappa_q'(0) = 0$. The zero function is not a descriptor covered by Theorem 1. It may be conceived as a limiting case corresponding to $q = \infty$.

4. HINTS TO THE LITERATURE

The formula (3.9) for a measure of entropy first appeared in the mathematical literature in Havrda and Charvát [6] and, independently, in Daróczy [5]. The latter author emphasized the characterization via functional equations, cf. also [1] and the more recent reference work [3]. The first appearance in the physical literature is due to Lindhard and Nielsen [11], where the property of *composability* – the ability to determine the entropy of a combined system from the entropies of its component subsystems – was the motivating principle. Subsequently, Lindhard gave a careful treatment of aspects of the measuring process, cf. [10].

The trend-setting publication [15] from 1988 by Tsallis marks the efficient promotion within the physical community of the new entropy measures. The paper triggered much research as also witnessed by the more than 2000 entries in the database maintained by Tsallis. At the time of publication, Tsallis was unaware of the earlier research. Regarding [11] and [10], these papers were largely unnoticed, probably due to their mathematical and somewhat lengthy style. However, there is a casual reference to Lindhard's work in one of Jaynes' papers, [8].

The success of Tsallis in launching the entropy measures which now bear his name is due to the direct approach and the fact that when combined with Jaynes *Maximum Entropy Principle*, cf. [7], main problems of statistical physics lead to *power laws*, a class of distributions which was and still is very popular as the basis for modelling when heavy-tailed distributions are involved. The present approach appears to be original, though inspired by and in line with earlier game theoretical considerations, cf. [14]. Because of a relation to Bregman divergences, we also point the reader to [12] and works referred to there.

FINAL REMARKS. The essence of our findings is that the family of Tsallis entropies can be derived based on two principles, the essential principle which allows for an interaction between truth, belief and experience and then a more innocent and

natural variational principle, that optimal performance is obtained when there is a perfect match between truth and belief.

It should be emphasized that though these principles may be viewed as axioms, they are intended as key elements of an interpretation behind the quantities they lead to, typically entropy, divergence and description cost.

Further research on the fundamental nature of the quantities characterized is much desired. In particular, we need to understand the mechanisms behind interaction and also, there is a need for a more complete interpretation of descriptors, ideally as clear and convincing as the coding interpretation of the classical quantities due to Shannon, cf. [13]. In this connection, Ahlswede [2] and references there may be relevant.

Acknowledgments. I thank Constantino Tsallis for comments resulting in a greater precision regarding terminology and past development, and Stig Stenstrup for pointing me to the important references [11] and [10].

References

[1] J. Aczél and Z. Daróczy. *On measures of information and their characterizations* (Academic Press, New York, 1975).

[2] R. Ahlswede."Identification Entropy". In Ahlswede et al, editor, *General Theory of Information Transfer and Combinatorics*, *Lecture Notes in Computer Science* **4123** 595–613, Springer, Berlin (2006).

[3] P. Sahoo B. Ebanks and W. Sander. *Characterizations of Information Measures* (World Scientific, Singapore, 1998).

[4] T. Cover and J. A. Thomas. *Elements of Information Theory* (Wiley, 1991).

[5] Z. Daróczy. Generalized Information Functions. *Information and Control* **16** 36–51 (1970).

[6] J. Havrda and F. Charvát. Quantification method of classification processes. Concept of structural a-entropy. *Kybernetika* **3** 30–35 (1967). Review by I. Csiszár in MR **34** (8875).

[7] E. T. Jaynes. Information theory and statistical mechanics, I and II. *Physical Reviews* **106** and **108** 620–630 and 171–190 (1957).

[8] E. T. Jaynes. Where do we Stand on Maximum Entropy? In R.D. Levine and M. Tribus, editors, *The Maximum Entrropy Formalism*, 1–104, M.I.T. Press, Cambridge, MA (1979).

[9] D. F. Kerridge. Inaccuracy and inference. *J. Roy. Stat. Soc. B.* **23** 184–194 (1961).

[10] J. Lindhard. On the Theory of Measurement and its Consequences in Statistical Dynamics. *Mat. Fys. Medd. Dan. Vid. Selsk.* **39** (1), 1–39 (1974).

[11] J. Lindhard and V. Nielsen. Studies in Statistical Dynamics. *Mat. Fys. Medd. Dan. Vid. Selsk.* **38** (9), 1–42 (1971).

[12] J. Naudts. Generalised exponential families and associated entropy functions. *Entropy* **10** 131–149 (2008).

[13] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.* **27** 379–423 and 623–656 (1948).

[14] F. Topsøe. Exponential Families and MaxEnt Calculations for Entropy Measures of Statistical Physics. In Qurati et al, editors, *Complexity, Metastability, and Non-Extensivity, CTNEXT07 AIP Conference Proceedings* **965** 104–113 (2007).

[15] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Physics* **52** 479–487 (1988). See http://tsallis.cat.cbpf.br/biblio.htm for a comprehensive and updated bibliography.

January 20, 2009